The 'big deal' approach to acquiring e-books: a usage-based study

This paper presents the results of an analysis of COUNTER e-book usage reports for SpringerLink e-book collections purchased since July 2008 at the University of Liverpool. The usage reports were augmented with contextual title-level information drawn from Springer's eBooks Title List. The combined data was used to study how usage of e-books is influenced by factors such as the subject area, the year of publication and the length of time since the collection was acquired. Analysis was performed to simulate the effects of user-driven purchasing to determine whether this model could apply to this type of content. The study concludes that this 'big deal' approach has worked well: all but one of the subject areas have been well used, the number of unused titles continues to diminish each year, older titles continue to attract significant usage, and the cost-per-use is relatively low. The challenge remains to find pricing models that allow more libraries to acquire e-book collections when budgets may be largely committed to journals.

TERRY BUCKNELL
Electronic Resources Manager
University of Liverpool

Introduction

The University of Liverpool Library's first significant foray into e-books occurred in summer 2006 with a subscription to ebrary Academic Complete. This provided a critical mass of titles and usage grew rapidly. User demand led to that collection being augmented through the acquisition of single-purchase titles through ebrary and later other e-book aggregators, dawsonera and MyiLibrary.

Whilst the ability to access books online was positively received, it became apparent that users disliked digital rights management (DRM) intensely. Feedback from library users was that they expected to be able to view, download and print e-book chapters just as they already did with e-journal articles: as DRM-free PDF files. As a result the library amended its strategy to increasingly buy e-books on publishers' own platforms. These e-books were often in bundles acquired at the end of the financial year if sufficient funds were available.

The Library first acquired Springer e-books, the complete collection for copyright years 2005 to 2008, at the end of the 2007/08 financial year in July 2008. The following July the complete 2009 collection was purchased. The timing of these purchases inevitably resulted in some duplication, as titles that had been bought in print or as single e-books through aggregators were later acquired in the collections. To avoid this double-purchasing, the library amended its strategy to purchase the 2010 collections at the end of 2009, mirroring the journal purchasing timetable. Any purchasing of Springer books outside of this arrangement is heavily discouraged. To help satisfy demand for print, Liverpool became one of the first libraries in Europe to join Springer's MyCopy programme[1]. MyCopy allows users to order any Springer e-book (with a few exceptions) as a personal print-on-demand soft cover print book for a heavily discounted price.

As part of its acquisition of the 2010 Springer e-book collections, University of Liverpool Library agreed to partner with Springer to study how e-books were being used and received, through a three-part project: analysis of usage reports, an extensive online user survey, and through focus groups. This article reports on the first part of this study – the analysis of usage reports.

Standard COUNTER BR2 reports were augmented with data from Springer's eBook Title List, using Excel's VLOOKUP function[2] to bring contextual title-level information, like copyright year and subject, into the usage reports.

About Springer e-books

Springer e-books are sold in 12 (formerly 13, but Birkhäuser Architecture has recently been sold) annual subject collections, with each year's collections containing around 3,500 titles. From 2005 forward all Springer book titles have been published online as well as in print. The collection is strong in STM subjects plus business and management, with a smaller number of titles in other humanities and social science subjects. The e-book collections are characterized by a high proportion of research monographs, contributed volumes and conference proceedings but also include textbooks and major reference works. They are mainly aimed at researchers from the advanced undergraduate level upwards

The SpringerLink platform presents e-book chapters alongside e-journal articles on a single interface. Each chapter is a PDF file that may be viewed, downloaded or printed in its entirety. MARC records are provided for each collection from 2005 onwards, and COUNTER-compliant usage reports are provided.

Results of e-book usage statistics study

Overall use of e-books

Figure 1 shows the pattern of use of Springer e-books at the University of Liverpool since January 2008, with e-journal usage included for context. The use of e-books after the first purchase in July 2008 is significant, with over 46,000 chapters downloaded from July 2008 to June 2009. This growth trend has continued from July 2009 forward and, based on figures measured to March 2010, would result in over 87,000 chapter downloads for the year ending July 2010 – an increase of 88% over the previous year.

It is notable that the use of Springer e-journals increased significantly between 2008 and 2009, even after excluding the usage of an additional collection that was added to Liverpool's big deal in 2009. This suggests that having access to e-books on the same platform as e-journals does have an inflationary effect on the usage of the e-journals. This will be closely watched to see if this pattern continues into 2010 and also if there are similar effects on other combined-content platforms.

Use of e-books by book type

Table 1 sets out the constituent parts of the e-book collections that were purchased, together with their respective e-book usage. The majority (35%) of the e-books in the collection are scholarly monographs and they attracted the highest volume of usage in absolute terms, with their percentage of downloads being roughly equal to their volume contribution. Usage of Proceedings was below their contribution in terms of downloads but this is perhaps because users are more likely to want just

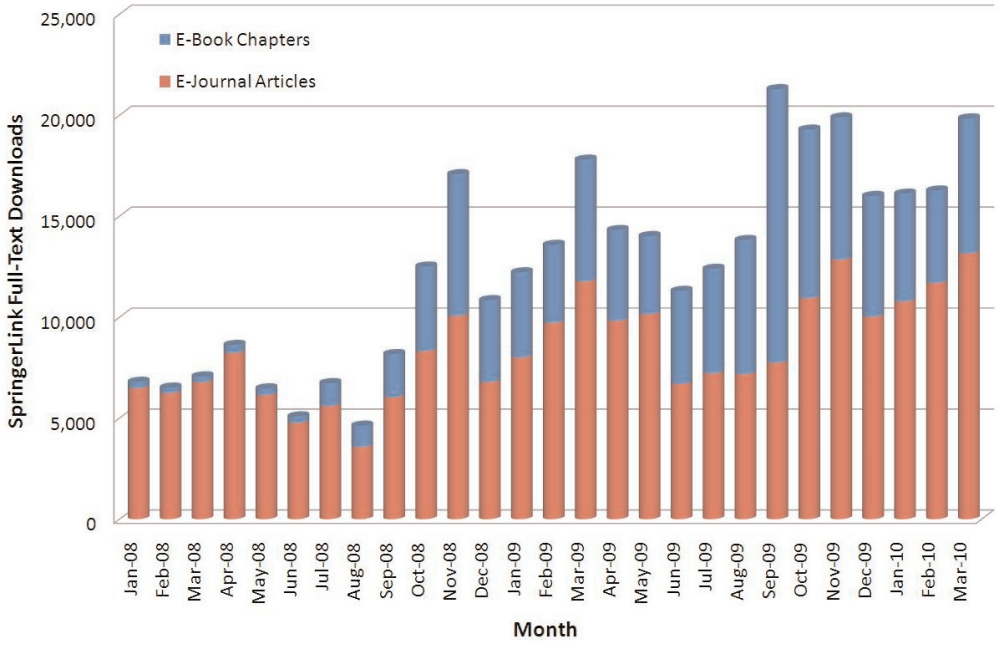

Figure 1. The University of Liverpool's monthly full-text downloads from SpringerLink from January 2008 onwards, broken down into e-journal article downloads and e-book chapter downloads. E-book usage prior to July 2008 is due to some freely available content.

a single article/chapter from this type of publication. Professional books and textbooks outperform their contribution which could be explained by the fact that these books generally have a broader appeal, especially at the undergraduate level. Reference works also significantly outperformed in terms of their % contribution to chapter downloads which can be explained by their large number of individual entries.

Year of publication

One major area of interest was the longevity of interest in scientific e-books: would they cease to be of much interest once they were more than a few years old? Table 2 shows that in 2009 the titles published in 2007 accounted for more total chapter downloads than any other year, and that the years 2006 and 2005 also contributed heavily to the overall usage. The 1997-2004 titles do not currently have MARC records available from Springer so this could account for their relatively lower usage. The downloads for 2009 titles are lower because this collection was not purchased until July 2009 and the collection was not completed until the end of 2009.

The pattern for the *proportion* of titles in each year's collection that were used in 2009 is even more uniform: four copyright years, 2005 to 2008, had over 45% of their individual titles accessed in 2009, Table 2 .

Figure 2 shows that once the 2009 collection had been purchased, its monthly share of the downloads quickly surpassed that of the earlier collections, even though that collection was not complete until the end of 2009. Nevertheless, titles from each of the previous years continued to make a significant and sustained contribution to the overall usage, much more so than older e-journals.

Differences between subject areas - overview

Figure 3 shows that different subject areas exhibit much consistency in the usage, with between 40% and 60% of titles used in all of the subject collections. The one exception was Mathematics and Statistics where fewer than 30% of titles were used. This mirrors the e-journals usage in mathematics at the University which is lower than most other subjects. This may be a reflection that mathematicians read less text-based material than those in other disciplines, still prefer to read from printed sources, or need to read older sources that are not available online.

Differences between subject areas – computer science and biomedical and life sciences

Computer science and biomedical and life sciences are two subjects that are generally perceived to have a heavy reliance on the most up-to-date research, generally based around journals and conference proceedings. These two subject

Type of book	No of titles in collection	% of titles in collection	No of downloads	% of downloads
Monograph	6,037	35.6%	21,184	32.1%
Proceedings	3,951	23.3%	11,928	18.1%
Contributed volume	2,519	14.9%	8,528	12.9%
Professional book	1,379	8.1%	7,585	11.5%
Textbooks	1,223	7.2%	7,701	11.7%
Handbook / Ref work / Encyclopedia	713	4.2%	6,463	9.8%
Other	1,132	6.7%	2,549	3.9%
Grand Total	**16,954**	**100.0%**	**65,938**	**100.0%**

Table 1. Make-up of Springer e-book collections by book type

Copyright Year	Chapters Downloaded in 2009	Percentage of Collection used in 2009
1997-2004	7,459	38.8%
2005	11,126	48.0%
2006	12,939	46.6%
2007	15,960	52.7%
2008	15,634	53.2%
2009	10,815	40.2%

Table 2. Usage of calendar year collections in 2009 showing the number of full-text chapter downloads and the proportion of each collection used

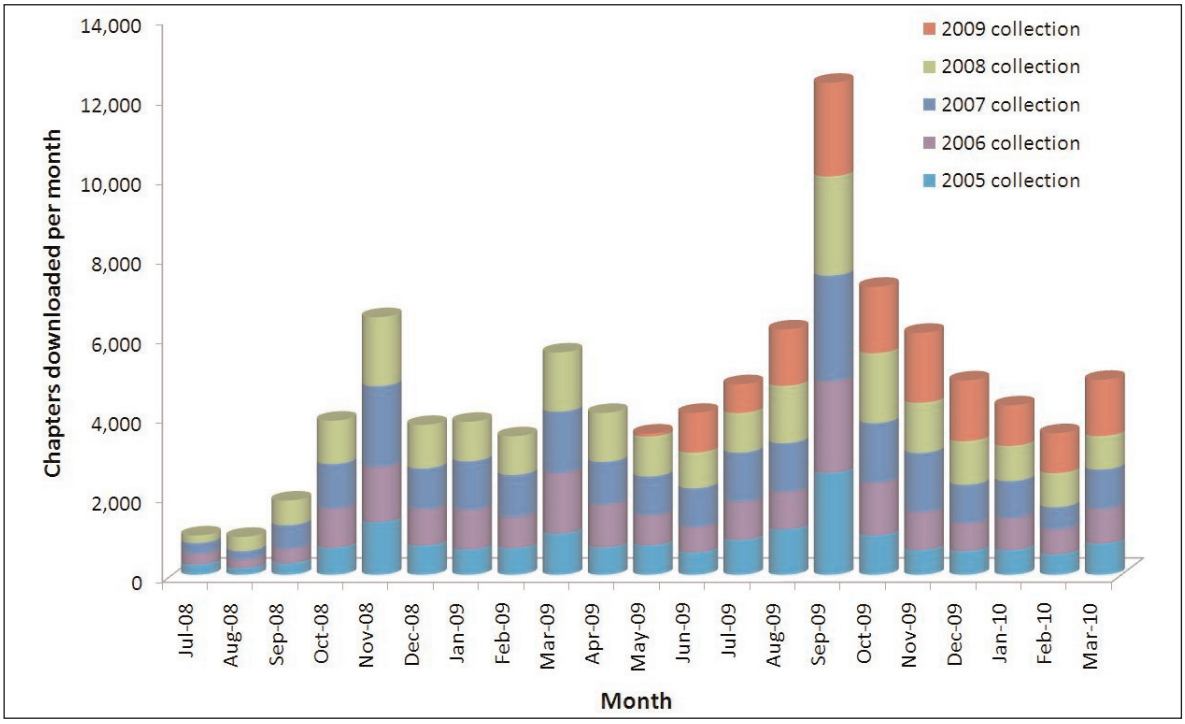

Figure 2. The University of Liverpool's monthly full-text e-book chapter downloads from SpringerLink from January 2008 onwards, broken down by calendar year collection

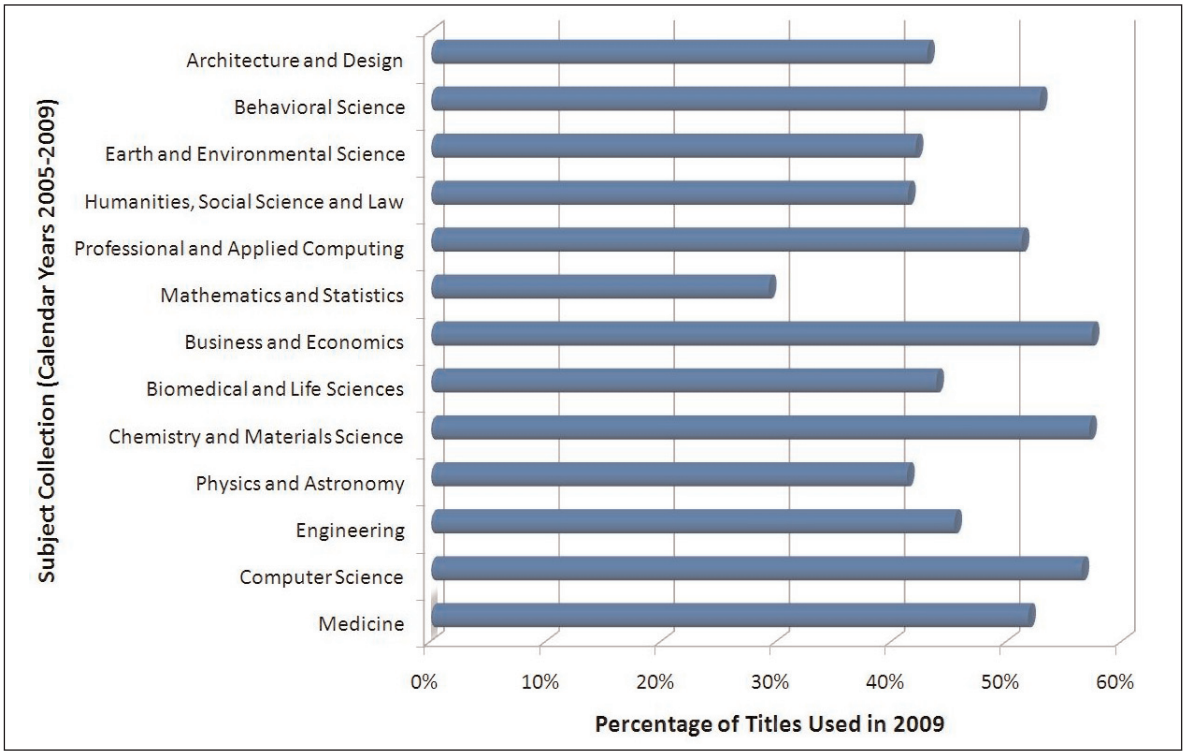

Figure 3. The proportion of titles in each subject collection (2005 to 2009 calendar years) that were used in full text in 2009

collections were investigated in more detail to see if this affected usage patterns. Figure 4 shows that even these subjects show little decline in the proportion of titles used in older collections. Figure 5 shows that fewer *chapters* were downloaded from older collections in Biomedical and Life Sciences but that there was little effect in Computer Science.

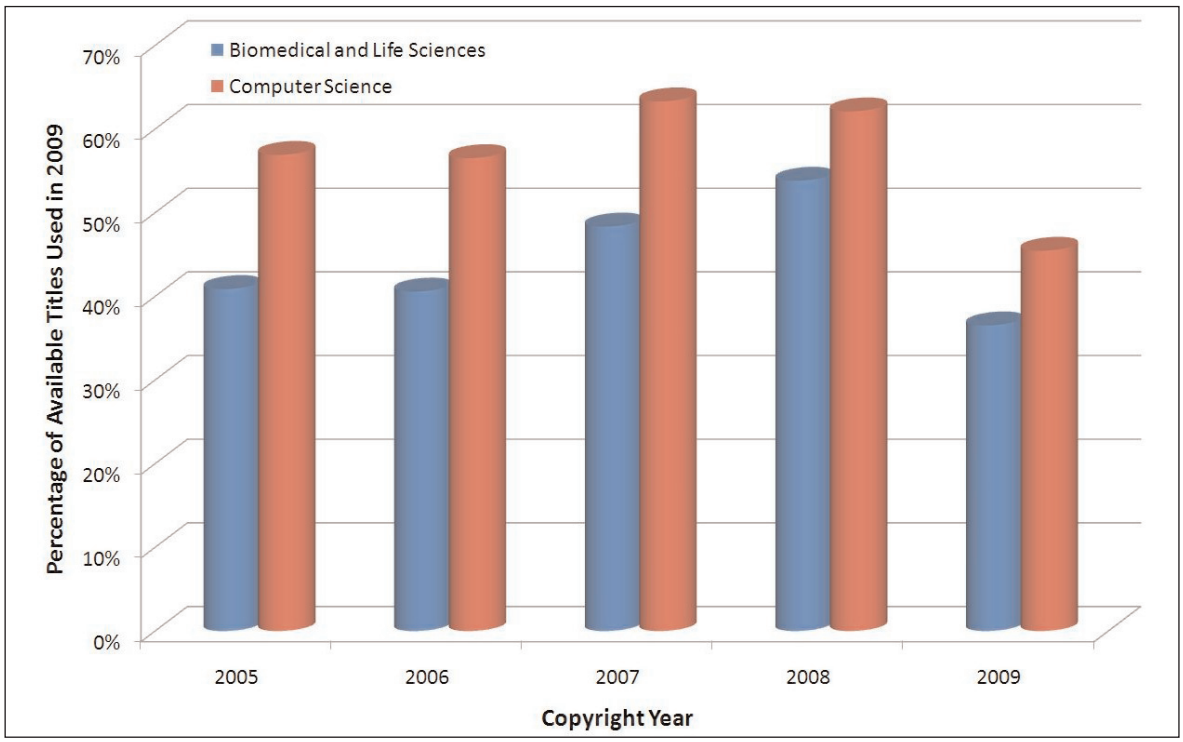

Figure 4. Full-text chapter downloads from the Biomedical and Life Sciences and Computer Science collections for the 2009 calendar year, broken down by copyright year.

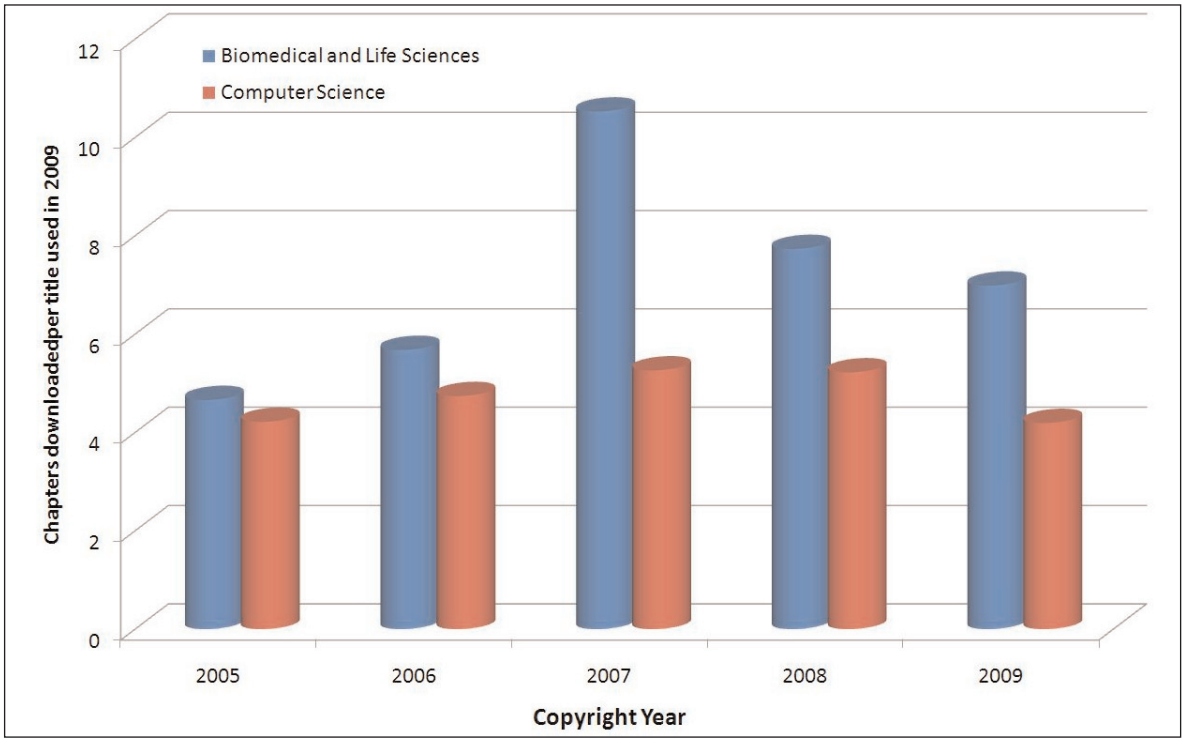

Figure 5. The proportion of Biomedical and Life Sciences and Computer Science titles in each calendar year collection that were used in full text in 2009.

Further analysis of e-book usage statistics

Effectiveness of book selection policy

The University of Liverpool Library acquires print books only when specifically requested by an academic department, either for teaching or research purposes, or when usage of existing copies demonstrates that additional copies are required. The library does not engage in approval plans and although some titles are acquired

through standing orders the use of these is declining.

Figure 6 shows the relationship between the total circulation figures for print books and the usage of the same titles in e-book format in 2009. Overall, there was little correlation between loans of print copies and their online usage. Indeed, if anything it appears that the titles that were borrowed most were used least online. This confirms that when users want to immerse themselves in a book thoroughly they prefer to read a print copy, but they are happy to use an online version if they just want to read a chapter or two. Nevertheless, the titles that had been bought in print performed slightly better online than the collection average. Of the titles that were held in print, 63% were used online in 2009, compared to 48% of the all titles in the collection. The library only owns 3.4% of the titles in the 2005-2009 collections, but they accounted for 4.4% of the titles that were used and 6.2% of the chapter downloads.

Overall in 2009 alone, 48% of all e-books from 2005 to 2009 were used at least once, which compares well against the widely-quoted figure of 40% of print books being unused six years after purchase[3]. In contrast, fewer than 40% of the purchased Springer e-books have not been used within *two* years of purchase.

Pattern of e-book usage from year to year

Several institutions and aggregators are currently experimenting with a user-driven purchasing model for e-books. This seems an attractive option for many libraries to ensure that every title that is purchased is used at least once. These models typically open up access to all titles for a limited period of time (either a fixed period or until the money runs out). At the end of the period, access to the non-purchased titles is switched off. Would this approach work if made available for Springer e-books?

Figure 7 shows the relationship between the number of chapters downloaded in the second half of 2008 and in the whole of 2009 for each title in the 2007 copyright year collection. This collection was chosen for analysis because it was complete, and owned, throughout the period under consideration. The usage pattern shows many titles that were well-used in 2008 received little or zero usage in 2009 and equally many titles from 2007 used heavily in 2009 were not used in 2008. In other words, past usage is not a good predictor of future usage: the titles selected during the limited period of a user-driven purchasing model are probably not the best titles to purchase for the long term.

To illustrate this further, only 11 of the 50 best-used titles in 2009 were also among the best-used

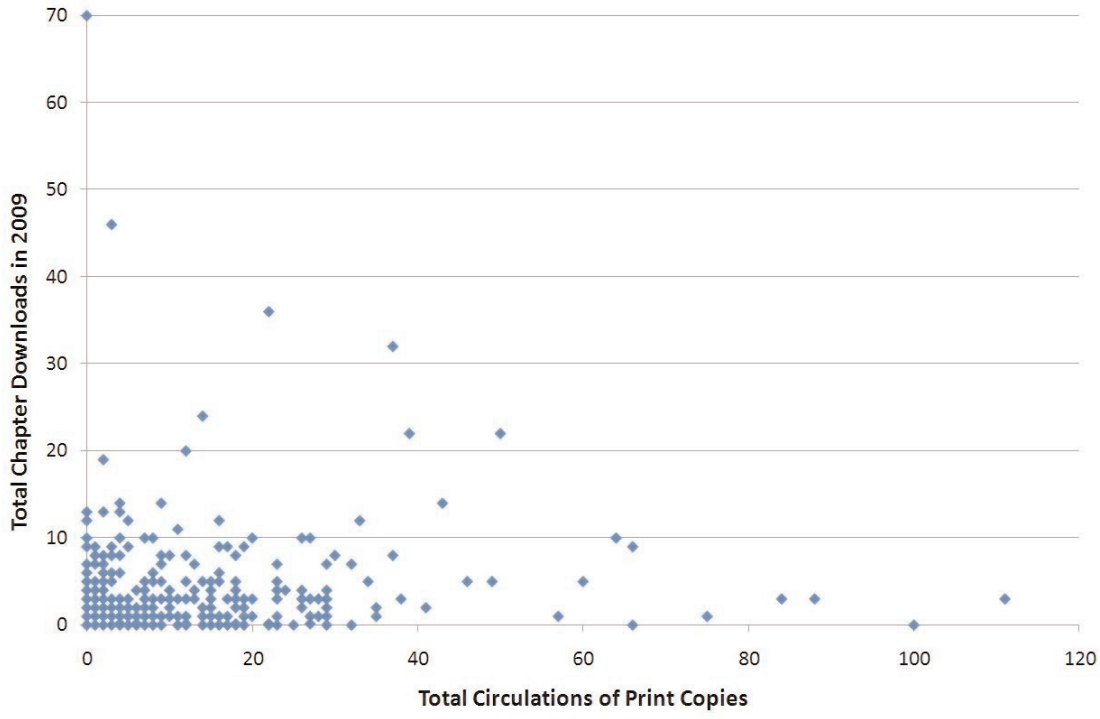

Figure 6. Relationship between e-book full-text chapter downloads in 2009 and total loans of print copies for titles owned in both formats. Each point on the graph represents a book.

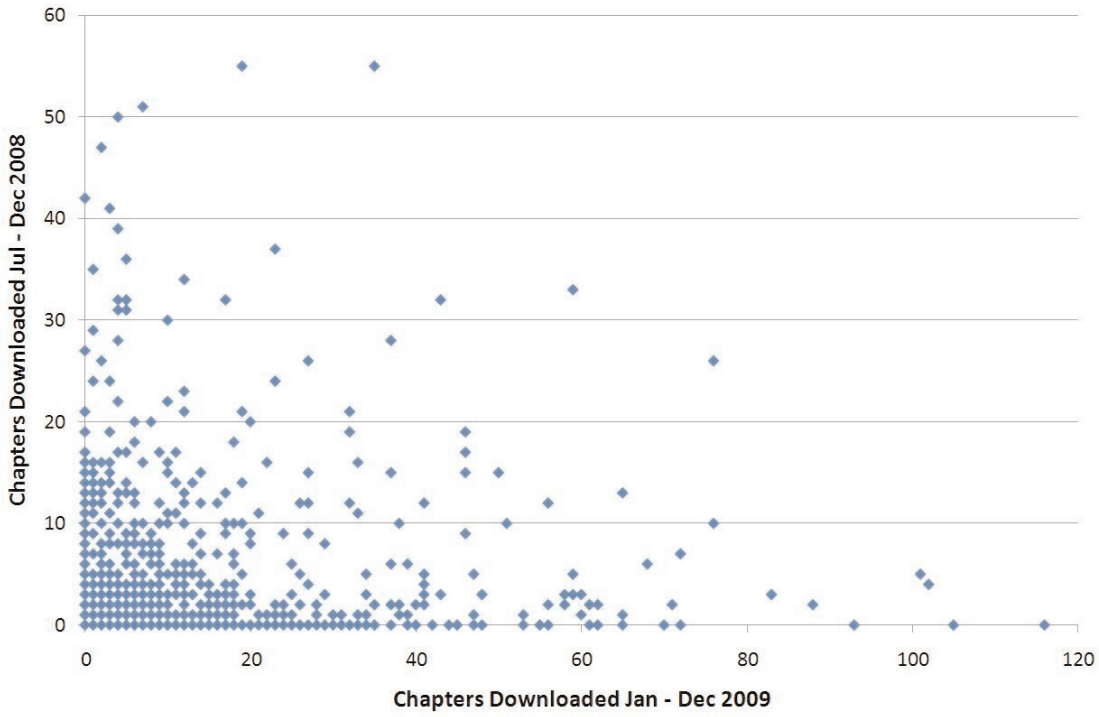

Figure 7. Relationship between e-book full-text chapter downloads in July to December 2008 and January to December 2009 for the 2007 copyright year collection (all subjects). Each point on the graph represents a book

titles in the *first quarter* of 2009. Thus if usage in the first quarter of 2009 had been used to select 50 titles for purchase, then 39 of the year's most heavily-used titles would have been missed, including the best-used title, and the fourth , fifth, eighth, ninth and tenth best-used titles. Equally, the library would have purchased some low-use titles – down to the 473rd best-used for the year.

Frequency of use

Analysis of the number of months for which each title was used showed that only one title had been used every month in 2009 and fewer than 5% of the titles had been used in five or more months of the year. More than 50% of titles were used between one and four times a year, and more than 45% of titles were not used at all (see Table 3). Note that the 'long tail' of infrequently used titles forms the bulk of the usage: titles used between one and four times in 2009 constituted 77.6% of the chapter downloads in 2009. This is *not* a case of a few high-use titles dominating usage of the collection: the breadth of the collection is key.

It should be noted that these figures just refer to usage in one calendar year. The longer a collection has been owned, the greater the proportion of the collection that will have been used. Figure 8 shows a remarkable consistency in the way that an increasing proportion of each collection is used over time.

About 12 months after acquisition, around 45% of the titles in each collection have been used; by 21 months over 60% have been used. It is quite remarkable to consider that 45 titles published in 2005 and purchased in July 2008 were used for the first time in March 2010 - and four of those had ten

No of Months Used	No of Titles	% of Collection	% of Downloads
0	8,810	46.3%	-
1	5,492	28.9%	26.9%
2	2,370	12.5%	22.8%
3	1,167	6.1%	17.9%
4	543	2.9%	10.0%
5–8	598	3.1%	18.5%
9–12	56	0.3%	3.9%

Table 3. Frequency with which titles in the 2005-2009 calendar year collections (all subjects) were used in 2009

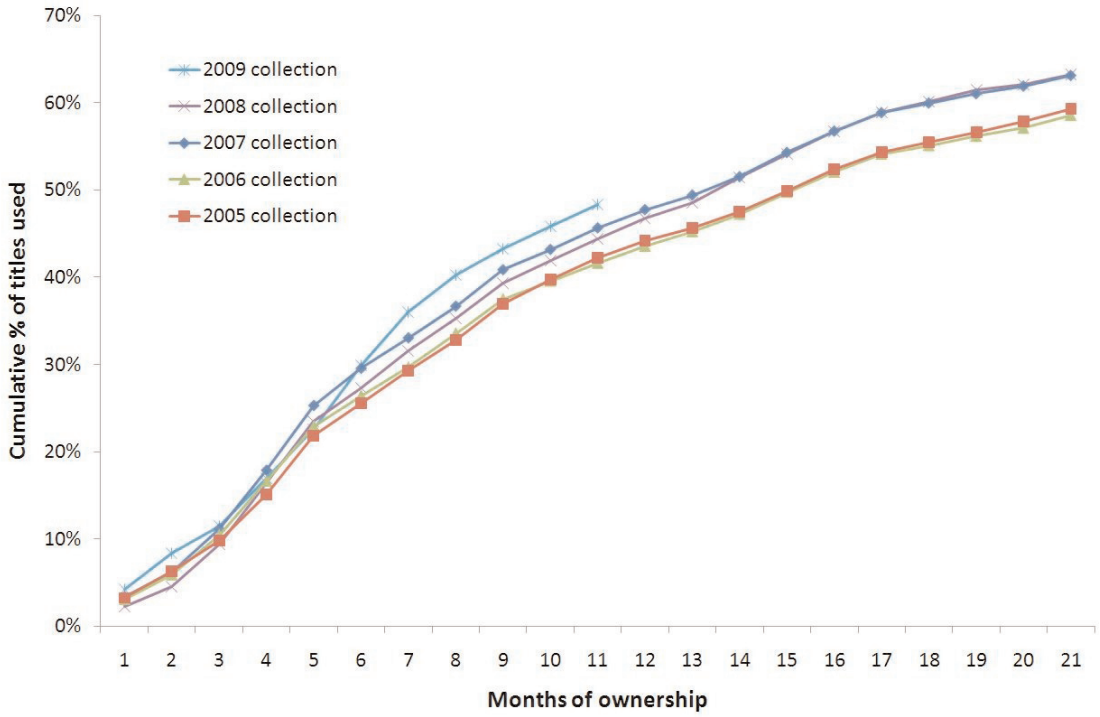

Figure 8. Cumulative percentage of titles within each calendar year collection (all subjects) that have been used in full text since the month of purchase. The 2005-2008 collections were purchased in July 2008; the 2009 collection in July 2009.

or more chapters downloaded in that month. It will be interesting to investigate whether collections purchased from other publishers exhibit the same shaped curve.

Measurement of e-book and e-book collection value

It is difficult to calculate and compare the value of e-book collections, in part because of the variability in how e-book platforms function and how they report usage[4]. COUNTER released its Code of Practice (CoP) for books and reference works in March 2006. This includes the COUNTER Book Report 2 (BR2) *Number of successful section requests by month and title* which has been implemented by many aggregators and publishers since. However, the COUNTER CoP defines a 'section' merely as 'a subdivision of a book or reference work'. This means that (depending on the platform) a dictionary definition, a single page, a group of five pages or a chapter can all count as a 'section'. Neither does the report state how a 'section' is defined for that platform; that is left to the librarian to know, or find out.

Cost per chapter as a metric of value

For Springer e-books, COUNTER BR2 counts full-text *chapter* downloads so a cost-per-chapter for the usage in 2009 can be calculated by dividing the price for the 2009 collections by the numbers of chapters downloaded in 2009. The University of Liverpool's total cost-per-download for Springer e-book chapters in 2009 was £1.43, which appears to be slightly better than that reported elsewhere for e-book packages[5]. This compares to a cost-per-article for Springer e-journals in 2009 of £1.94, so the e-books look like good value. Value calculations based on aggregator platforms necessarily tend to be quoted as cost-per-page viewed. Assuming an average book chapter is around 20 pages long, then Liverpool's cost-per-page viewed for Springer e-books is about 7p (assuming all pages in a downloaded chapter are read).

To be absolutely correct, only the usage of the content published in 2009 should be used in the calculation as this is what the 2009 fee is paying for. Likewise, it should be the usage counted in every year from 2009 onwards that should be included in the calculation. This would mean it would take until the end of time to calculate the true value of the purchase! However, these arguments apply equally to e-journals where it is the norm to just divide the year's payment by that year's usage, so it seems justified to apply that same simplified calculation to e-books purchased in an annual collection.

Conclusions

Through fairly simple analysis of COUNTER reports it has been shown that acquiring Springer e-books as a big deal has been a good value investment for the University of Liverpool. Each subject area has been well used (with the usual exception of mathematics), the number of unused titles continues to diminish each year, and older titles continue to attract significant usage.

Acquiring a 'database of book chapters' seems to be the best approach for research-level e-books (research monographs, contributed works, conference proceedings, etc.). Coupled with Springer's MyCopy programme, this approach could help shift the library to a 'primarily electronic' book acquisitions policy for this type of material.

The study has highlighted that STM students and researchers make good use of book material even though STM faculty prefer to allocate their library budgets almost entirely to journals. At the University of Liverpool this leaves little scope for finding funds to acquire e-book collections, except through end-of-financial-year purchases. As more publishers start producing bundles of e-book content, perhaps with a view to bundling this with e-journal content, they need to understand that book budgets are often dwarfed by journal budgets and they may need to fundamentally rethink their pricing models so that libraries' book budgets can accommodate such offerings.

References

1. Springer's MyCopy: http://www.springer.com/librarians/e-content/my+copy?SGWID=0-165802-0-0-0 (accessed 30 April 2010).
2. Using the VLOOKUP Function in Excel 2007: http://msdn.microsoft.com/en-us/library/dd797422.aspx (accessed 30 April 2010).
3. Kent, A *et al*, *Use of Library Materials: The University of Pittsburgh Study*, 1979, New York, Marcel Dekker Inc.
4. Cox, J, Making Sense of E-book Usage Data, *The Acquisitions Librarian*, 2008, 19(3), 193–212.
5. Sprague, N and Hunter, B, Assessing e-books: Taking a closer look at e-book statistics, *Library Collections, Acquisition and Technical Services*, 2008, 32(3-4), 150–157.

■ **Terry Bucknell**
Electronic Resources Manager
Sydney Jones Library
University of Liverpool
Chatham St, PO Box 123
Liverpool, L69 3DA, UK
E-mail: t.d.bucknell@liverpool.ac.uk
Tel: +44 (0)151 794 2692
Fax: +44 (0)151 794 2681